Microsoft Windows 2000 Professional Reliability:

A comparative study of the reliability of Windows 2000 Professional
Microsoft Windows 98
And Microsoft Windows NT Workstation 4.0

ZDLABS

January 2000

Disclaimer of Warranties; Limitation of Liability:

ZD LABS HAS MADE REASONABLE EFFORTS TO ENSURE THE ACCURACY AND VALIDITY OF ITS TESTING, HOWEVER, ZD LABS SPECIFICALLY DISCLAIMS ANY WARRANTY, EXPRESSED OR IMPLIED, RELATING TO THE TEST RESULTS AND ANALYSIS, THEIR ACCURACY, COMPLETENESS OR QUALITY, INCLUDING ANY IMPLIED WARRANTY OF FITNESS FOR ANY PARTICULAR PURPOSE. ALL PERSONS OR ENTITIES RELYING ON THE RESULTS OF ANY TESTING DO SO AT THEIR OWN RISK, AND AGREE THAT ZD LABS, ITS EMPLOYEES AND ITS SUBCONTRACTORS SHALL HAVE NO LIABILITY WHATSOEVER FROM ANY CLAIM OF LOSS OR DAMAGE ON ACCOUNT OF ANY ALLEGED ERROR OR DEFECT IN ANY TESTING PROCEDURE OR RESULT.

IN NO EVENT SHALL ZD LABS BE LIABLE FOR INDIRECT, SPECIAL, INCIDENTAL, OR CONSEQUENTIAL DAMAGES IN CONNECTION WITH ITS TESTING, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. IN NO EVENT SHALL ZD LABS' LIABILITY, INCLUDING FOR DIRECT DAMAGES, EXCEED THE AMOUNTS PAID IN CONNECTION WITH ZD LABS' TESTING. CUSTOMER'S SOLE AND EXCLUSIVE REMEDIES ARE AS SET FORTH HEREIN.

Origin of report

ZD Labs prepared this report under contract from Microsoft Corporation.

Table of Contents

1.0 Executive Summary

Microsoft Corporation commissioned ZD Labs to compare the reliability of Windows 2000 Professional operating system with that of Windows 98 Second Edition (SE) and Windows NT Workstation 4.0 with Service Pack 6a (SP6a). In the absence of a standard desktop reliability benchmark, Microsoft requested that ZD Labs create a custom independent test.

We designed the reliability test to measure the average length of time that an operating system could perform common tasks before needing to be rebooted. We developed a Rational Visual Test script that automated these common tasks: browsing the Web, editing documents, working with spreadsheets, and accessing data in a database. The automated scripts forced the computers under test to do a lot more work than a typical user would. The test script looped continuously, moving from one task to another until it encountered a problem.

We ran the test script on each operating system for thirty consecutive days and measured the total amount of time that our test stressed the system, i.e. the time spent looping the script, and logged the total number of reboots during the thirty-day test period. (This form of measurement takes into account the problem that a computer might have been stuck overnight before it was restarted.) From these two numbers, we calculated the average time each operating system could perform real work before needing to be rebooted (average uptime). To try and make these numbers more understandable, we used the number of eight-hour workdays as the unit of measure.

During the duration of this test, Windows 2000 Professional experienced no errors and never needed to be rebooted. There are a number of reasons why an operating system could need rebooting: the operating system could lock-up or crash, what we refer to in this report as a hard error, or the operating system could become unstable or leak enough resources so that user applications no longer run as they should. We referred to these errors as soft errors. Both Windows NT Workstation 4.0 and Windows 98 SE required reboots. Windows NT Workstation 4.0 experienced only soft errors, while Windows 98 SE experienced a mix of soft errors and hard errors.

The graph below gives a better perspective of Windows 2000 Professional reliability. At the end of thirty calendar days (ninety eight-hour workdays), Windows 2000 had not encountered a single error. We do not know how long Windows 2000 might have run if the test had been continued.

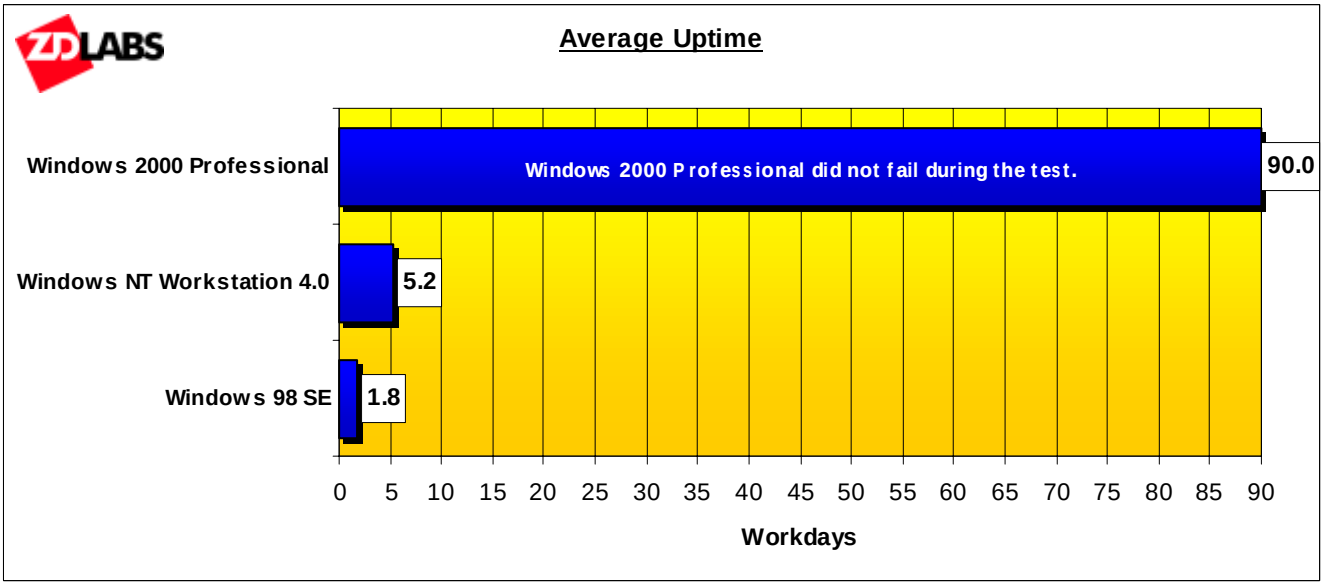

Figure 1: Average uptime in eight-hour workdays.

While the experiences of individual users may vary from our test results, the reliability of Windows 2000 Professional was outstanding. It performed continuously and flawlessly for more than ninety business workdays without a single failure. Not only did it not encounter any problems during this extended period of testing, but also the amount of work done was considerably more than that of a typical user.

Although testing with different applications might show different results, based on our testing we conclude that the reliability of Windows 2000 Professional far exceeds that of Windows 98 SE and Windows NT Workstation 4.0.

2.0 Test Methodology

In this section we explain our test methodology for determining reliability. We describe how we measured reliability and give details of the test setup and execution.

2.1 Measuring Reliability

We designed the reliability test to measure the average length of time that an operating system could perform common tasks before needing to be rebooted. We developed a Rational Visual Test script that automated these common tasks: browsing the Web, editing documents, working with spreadsheets, and accessing data in a database. The automated scripts forced the computers under test to do a lot more work than a typical user would. The test script looped continuously, moving from one task to another until it encountered a problem.

While developing the reliability test, we found that it was not always easy to know when the operating system needed to be rebooted. We encountered many errors, but only a handful of those errors caused the operating system to fail in an obvious way, i.e. to lock-up or crash. So, we had to determine when an error was severe enough to merit rebooting the operating system. Making this determination turned out to be easier than we had at first thought.

We started by classifying errors as either soft errors or hard errors. Hard errors were non-recoverable errors that caused the operating system to lock-up or crash, while soft errors were errors that caused a single test loop to fail but that did not crash the operating system. Hard errors were obvious operating system failures, requiring either a Ctrl-Alt-Del or power reset to regain control of the operating system. Soft errors, however, were not obvious operating system failures.

After careful study, we adopted the “second chance” rule as our arbiter for deciding when to reboot the operating system in response to a soft error. The second chance rule was simple; when a test loop encountered a soft error we always attempted the test loop a second time. If the second attempt failed, resulting in two consecutive failed loops of the script, we rebooted the operating system before continuing the test.

The second-chance approach worked well for deciding when to reboot. We discovered that an operating system was often capable of completing further loops of the script after encountering a soft error. After two consecutive loops of the script failed, however, we found that the operating system was unable to continue the test unless we rebooted the system.

We ran the test script on each operating system for thirty consecutive days and measured the total amount of time that our test stressed the system, i.e. the time spent looping the script, and logged the total number of reboots during the thirty-day test period. (This form of measurement takes into account the problem that a computer might have been stuck overnight before it was restarted.) From these two numbers, we calculated the average time each operating system could perform real work before needing to be rebooted. To try and make these numbers more understandable, we used the number of eight-hour workdays as the unit of measure.

2.2 Test Setup

We conducted our reliability test on six identical Dell OptiPlex GX110 systems. All systems arrived directly from Dell, pre-loaded with each operating system. Except for installing the test applications, changing the screen resolution from 800x600x16bpp to 1024x720x16bpp and disabling all energy saving features, we made no changes to the configurations of the test systems.

We divided the six systems in to two matching test beds of three systems each. We conducted parallel tests on both test beds. Had one or more systems on the primary test bed encountered hardware failures, the second test bed would have provided backup test data. We experienced no problems with the Dell OptiPlex GX110s, however, and had no need to pull results from the backup systems.

One of the goals of the test was to stress as much of the operating system as possible. So, in addition to generating hard disk and video activity, we included CD-ROM, file server, and Web browsing activity to the workload mix. Word and Excel loaded original documents and spreadsheets from a CD-ROM and saved iterative versions of the files to a working directory on a server. We created six identical data CDs for the CD-ROM activity and attached all test clients to a server that provided file and Web services.

Windows NT 4.0 Server with Service Pack 5 and Internet Information Server (IIS) provided all of our file and Web services. The Web browsing portion of our test required the additional installations of i-Bench 1.0 Server, Windows Media Services and SQL Server 6.5. We configured the server as a primary domain controller and all test clients logged into a common domain account.

2.3 Visual Test Script

We developed a Rational Visual Test script that automated these common tasks: browsing the Web, editing documents, working with spreadsheets, and accessing data in a database. The automated script forced the computers under test to do a lot more work than a typical user would. The test script looped continuously, moving from one task to another until it encountered a problem.

The test script generated Web browsing activity by running four of ZD's i-Bench 1.0 performance tests: Load Simple Pages, Load Complex Pages, Java VM Processor test, and the 56Kb Windows Media Streaming Video test. i-Bench is a comprehensive, cross-platform benchmark that tests the performance and capability of Web clients as they take on the latest Web technology and features. The i-Bench tests provided an excellent mix of Web browsing activity.

For document, spreadsheet, and database activity the scripts ran the latest Microsoft Office 2000 applications: Word, Excel, and Access. We utilized our "hot-spot" research to generate a stressful workload. Hot spots are where demanding users tend to have to wait on their PC to complete a task.

Figure 2 and Figure 3 provide statistics for the documents used in the reliability test.

Word Document	Pages	Words	Characters	Paragraphs	Lines
Document One	4	3309	19158	23	198
Document Two	9	1930	10077	215	323

Figure 2. Microsoft Word documents

Excel Workbook	Worksheets	Data Cells
Workbook 1	3	32118
Workbook 2	1	378
Workbook 3	2	634
Workbook 4	6	1096

Figure 3. Microsoft Excel spreadsheets

The following outline details the tasks performed by a single loop of our Rational Visual Test script:

- Launched Internet Explorer, Microsoft Excel and Microsoft Word.
- Switched to Internet Explorer.
- Ran the Load Complex Pages Test.
- Ran the Load Simple Pages Test.
- Ran the Java Virtual Machine Processor Test.
- Ran the 56Kb Streaming Video Test.
- Switched to Microsoft Word and loaded Document One (37 KB) from the CD-ROM.
- Saved a working copy of Document One (37 KB) to the server work directory.
- Performed spelling and grammar check on Document One.

- Replaced select words with synonyms.
- Reformatted document text.
- Copied and pasted portions of the document text.
- Saved and closed Document One (40 KB).
- Loaded Document Two (192 KB) from the CD-ROM.
- Saved a working copy of Document Two (192 KB) to the server work directory.
- Inserted images, BARRELA.JPG (5 KB) and CHMPBUKT.JPG (8 KB) from the CD-ROM into the document.
- Performed a print preview of the entire document.
- Executed an macro to create a footnote in the document.
- Saved and closed Document Two (209 KB) to the server work directory.
- Switched to Excel.
- Loaded Workbook One (314 KB) from the CD-ROM.
- Saved a working copy of Worksheet One (314 KB) to the server work directory.
- Copied and pasted formulas into cells.
- Selected a range of columns and sorted them based on the values of a single column.
- Created a chart based on a range of selected cells.
- Created a second chart based on another range of selected cells.
- Moved among multiple worksheets in the workbook.
- Converted a selected range of data to a Microsoft Access database (416 KB) and linked the data in the database to the data in the worksheet.
- Saved and closed Workbook One (419 KB).
- Loaded Workbook Two (67 KB) from the CD-ROM.
- Saved a working copy of Workbook Two (67 KB) to the server work directory.
- Updated data values and recalculated the workbook.
- Performed a print preview of a single worksheet.
- Saved and closed Workbook Two (66 KB).
- Loaded Workbook Three (26 KB) from the CD-ROM.
- Saved a working copy of Workbook Three (26 KB) to the server work directory.
- Reformatted a range of cells.
- Performed a print preview of the entire workbook.
- Executed a macro that creates a chart in the worksheet.
- Performed a print preview of the chart.
- Saved and closed Workbook Three (30 KB).
- Reopened Workbook three (30 KB) from the server work directory.
- Loaded Workbook Four (22 KB) from the CD-ROM.
- Saved a working copy of Workbook Four (22 KB) to the server work directory.
- Converted a selected range of data to a Microsoft Access database (112 KB) and link the data in the database to the data in the worksheet.
- Inserted formulas into a worksheet that referenced cells in a different Workbook.
- Reformatted a range of cells in the worksheet.
- Saved and closed the Workbook Four (22 KB).
- Exited each of the three Applications.

2.4 Summary of Script I/O Activity

Figure 4 summarizes the type and amount of I/O generated by a single loop of the test script and the average amount of I/O generated during one hour of test uptime. The average I/O per hour was computed based on an average loop time of eight minutes.

Transaction Type	Read (KB)	Write (KB)	Total I/O (KB)	Average I/O per Hour (KB)
CD-ROM	671	0	671	5,033
File Server	30	1,972	2,002	15,015
Web Downloads	2,451	0	2,451	18,383
All	**3,152**	**1,972**	**5,124**	**38,430**

Figure 4. Summary of Script I/O Activity

Although we did not specifically include hard disk activity in the script, hard disk I/O obviously occurred during each loop of the script as a result of applications being launched, DLLs being loaded, caching of Web files, system registry updates, system page file activity, etc.

3.0 Findings and Results

We ran the test script on each operating system for thirty consecutive days and measured the total amount of time that our test stressed the system, i.e. the time spent looping the script, and logged the total number of reboots during the thirty-day test period. (This form of measurement takes into account the problem that a computer might have been stuck overnight before it was restarted.) From these two numbers, we calculated the average time each operating system could perform real work before needing to be rebooted. To try and make these numbers more understandable, we used the number of eight-hour workdays as the unit of measure.

During the duration of this test, Windows 2000 Professional experienced no errors and never needed to be rebooted. There are a number of reasons why an operating system could need rebooting: the operating system could lock-up or crash, what we refer to in this report as a hard error, or the operating system could become unstable or leak enough resources so that user applications no longer run as they should. We referred to these errors as soft errors. Both Windows NT Workstation 4.0 and Windows 98 SE required reboots. Windows NT Workstation 4.0 experienced only soft errors, while Windows 98 SE experienced a mix of soft errors and hard errors. Figure 5 summarizes the final results of our test. See Appendix B for raw results.

Operating System	Uptime (hours)	Soft Errors	Hard Errors	Reboots	Avg. Uptime (hours)	Avg. Uptime (workdays)
Windows 2000 Professional	720.00	0	0	0	720.00	90.0
Windows NT Workstation 4.0	538.99	13	0	13	41.46	5.2
Windows 98 SE	262.63	15	3	18	14.59	1.8

Figure 5: Summary of reliability test results.

The graph below gives a better perspective of Windows 2000 Professional reliability. At the end of thirty calendar days (ninety eight-hour workdays), Windows 2000 had not encountered a single error. We do not know how long Windows 2000 might have run if the test had been continued.

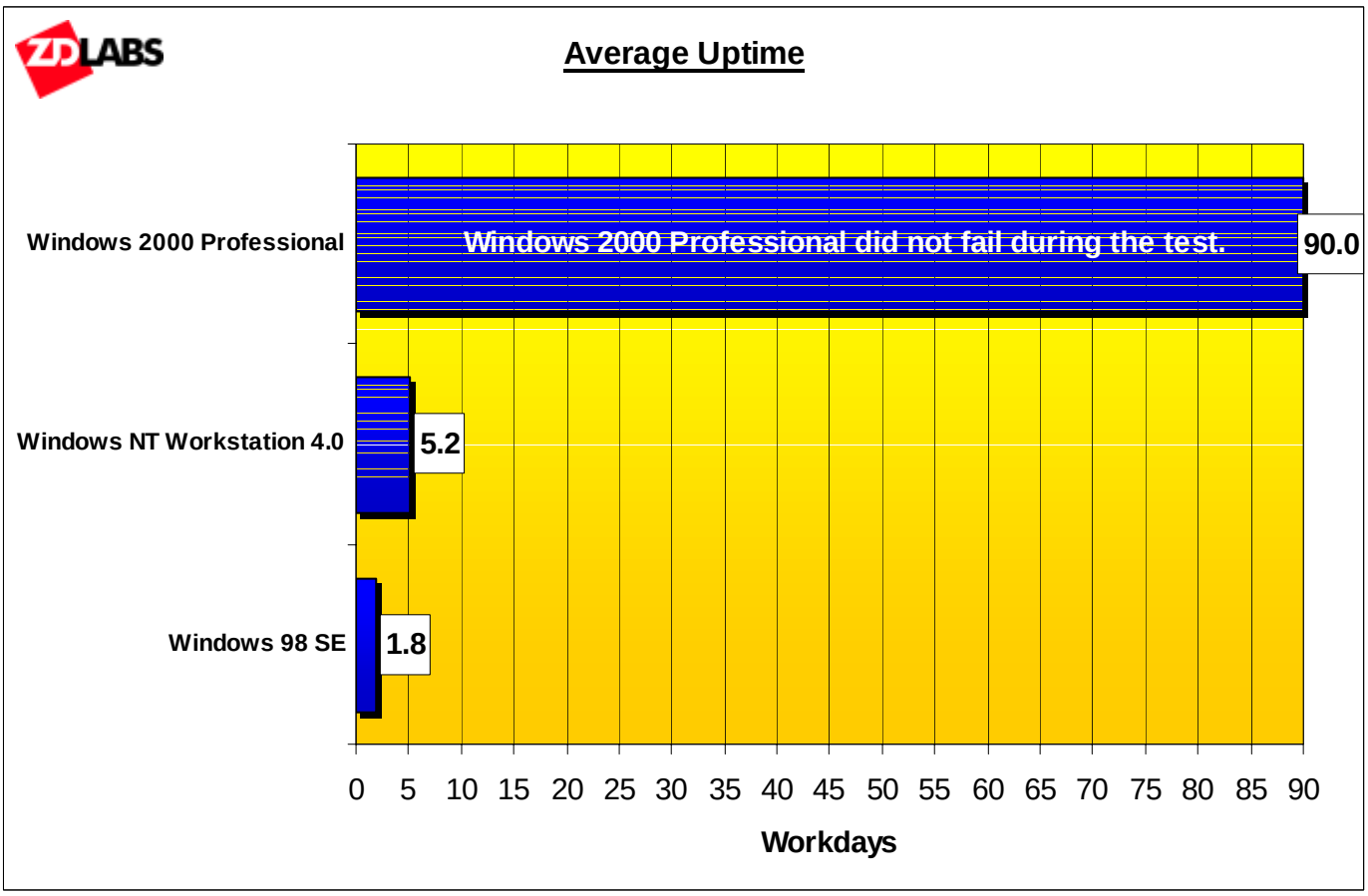

Figure 6: Average uptime in eight-hour workdays.

While the experiences of individual users may vary from our test results, the reliability of Windows 2000 Professional was outstanding. It performed continuously and flawlessly for more than ninety business workdays without a single failure. Not only did it not encounter any problems during this extended period of testing, but also the amount of work done was considerably more than that of a typical user (refer to Section 2.0 Test Methodology for details regarding the type and amount of work performed).

Although testing with different applications might show different results, based on our testing we conclude that the reliability of Windows 2000 Professional far exceeds that of Windows 98 SE and Windows NT Workstation 4.0.

Appendix

A. System Disclosure

Test systems	
Machine Description	Dell OptiPlex GX110
Processor	Pentium III
Number of Processors	1
Processor Speed	600 MHz/133 MHz front side bus
L1 Cache	32 KB
L2 Cache	512 KB
System RAM	128 MB
BIOS	Phoenix ROM BIOS Plus v1.10A01
HD Model	Western Digital WD136AA
HD Size	13.6 GB
HD Controller	Intel 82801AA Ultra ATA
Windows 2000 File System	FAT32 / DMA
Windows 98 SE File System	FAT32 / DMA
Windows NT 4.0 File System	FAT16 / PIO
Video Adapter	Intel 810e integrated
Video Memory	4 MB
Windows 2000 Video Driver	I81xnt5.sys v5.11.0133.3; i81xdnt5.dll
Windows 98 SE Video Driver	I81x.vxd v4.11.01.1361
Windows NT 4.0 Video Driver	I81xnt4.sys v4.03.1381.1345; i81xdnt4.dll v4.0.32
Video Resolution Assigned	1024 x 768
Color Depth Assigned	16 bpp
Sound Board	SoundBlaster AWE64 PCI
NIC	Integrated 3Com Etherlink 10/100 PCI LAN
Printer Driver (NUL port)	HP LaserJet 5P/MP PostScript
CD-ROM Manufacturer	Lite-On LTN483s
CD-ROM Speed	20X-48X Max.
Windows Version	Windows 2000, Build 2195
	Windows 98, Build 2222
	Windows NT 4.0, Build 1381, Service Pack 6a

Figure 7: System disclosure information

B. Test Results

Time of Reboot	Soft Errors	Hard Errors	Runs	Uptime
Did not fail during test.	0	0	5585	30 days

Figure 8. Windows 2000 Professional Results

Time of Reboot	Soft Errors	Hard Errors	Runs	Uptime (ms)
Sat Jan 08 11:15:54 2000	1		326	149105904
Mon Jan 10 08:47:38 2000	1		326	148910272
Wed Jan 12 03:40:49 2000	1		325	148663400
Fri Jan 14 03:16:00 2000	1		327	149594939
Sun Jan 16 05:31:23 2000	1		326	149089882
Wed Jan 19 06:05:26 2000	1		327	149435778
Fri Jan 21 06:24:57 2000	1		326	149060022
Sun Jan 23 17:03:05 2000	1		328	150925704
Wed Jan 26 03:57:38 2000	1		326	148928968
Sat Jan 29 06:09:50 2000	1		327	149350297
Mon Jan 31 10:07:07 2000	1		326	148886443
Thu Feb 03 04:19:25 2000	1		327	149391424
Sat Feb 05 13:05:42 2000	1		326	139255904
	13	0	4243	1940378937

Figure 9. Windows NT Workstation 4.0 Results

Time of Reboot	Soft Errors	Hard Errors	Runs	Uptime (ms)
Tue Jan 04 07:37:21 2000	1		116	57778088
Tue Jan 04 23:42:48 2000		1	86	40320211
Thu Jan 06 01:46:59 2000	1		117	54982682
Fri Jan 07 00:10:39 2000	1		81	37970342
Sat Jan 08 03:06:31 2000	1		115	54033508
Mon Jan 10 21:47:23 2000	1		117	54860334
Thu Jan 13 02:59:44 2000	1		113	53117746
Fri Jan 14 04:12:40 2000	1		120	56195266
Sat Jan 15 04:14:13 2000	1		116	54436999
Tue Jan 18 04:50:08 2000	1		116	54500269
Thu Jan 20 05:49:05 2000		1	118	55189371
Sat Jan 22 02:27:36 2000	1		120	56297427
Tue Jan 25 01:45:09 2000	1		115	53962077
Fri Jan 28 04:47:50 2000	1		117	54760683
Sat Jan 29 06:14:36 2000	1		120	56186847
Sun Jan 30 09:26:11 2000	1		119	55715489
Wed Feb 02 03:16:40 2000	1		118	55261783
Wed Feb 02 22:25:47 2000		1	85	39903947
	15	3	2009	945473069

Figure 10. Windows 98 SE Results

1001 Aviation Parkway, Suite 400
Morrisville, NC 27560
Tel 919-380-2800
FAX 919-380-2899
www.zdlabs.com